

# ON NORMAL DOMINATION OF (SUPER)MARTINGALES

IOSIF PINELIS

ABSTRACT. Let  $(S_0, S_1, \dots)$  be a supermartingale relative to a non-decreasing sequence of  $\sigma$ -algebras  $(H_{\leq 0}, H_{\leq 1}, \dots)$ , with  $S_0 \leq 0$  almost surely (a.s.) and differences  $X_i := S_i - S_{i-1}$ . Suppose that for every  $i = 1, 2, \dots$  there exist  $H_{\leq (i-1)}$ -measurable r.v.'s  $C_{i-1}$  and  $D_{i-1}$  and a positive real number  $s_i$  such that  $C_{i-1} \leq X_i \leq D_{i-1}$  and  $D_{i-1} - C_{i-1} \leq 2s_i$  a.s. Then for all real  $t$  and natural  $n$

$$\mathbf{E}f_t(S_n) \leq \mathbf{E}f_t(sZ),$$

where  $f_t(x) := \max(0, x - t)^5$ ,  $s := \sqrt{s_1^2 + \dots + s_n^2}$ , and  $Z \sim N(0, 1)$ . In particular, this implies

$$\mathbf{P}(S_n \geq x) \leq c_{5,0} \mathbf{P}(Z \geq x/s) \quad \forall x \in \mathbb{R},$$

where  $c_{5,0} = 5!(e/5)^5 = 5.699\dots$ . Results for  $\max_{0 \leq k \leq n} S_k$  in place of  $S_n$  and for concentration of measure also follow.

## CONTENTS

1. Introduction	1
2. Domination by normal moments and tails	4
3. Maximal inequalities	6
4. Concentration inequalities for separately Lipschitz functions	8
5. Proofs	11
5.1. Proofs for Section 2	11
5.2. Proofs for Section 3	14
5.3. Proofs for Section 4	17
References	18

## 1. INTRODUCTION

The sharp form,

$$(1.1) \quad \mathbf{E}f(\varepsilon_1 a_1 + \dots + \varepsilon_n a_n) \leq \mathbf{E}f(Z),$$

---

*Date:* February 2, 2008.

2000 *Mathematics Subject Classification.* Primary: 60E15, 60J65; Secondary: 60E05, 60G15, 60G50, 60J30.

of Khinchin's inequality for  $f(x) = |x|^p$  for the normalized Rademacher sum  $\varepsilon_1 a_1 + \dots + \varepsilon_n a_n$ , with

$$a_1^2 + \dots + a_n^2 = 1,$$

was proved by Whittle (1960) [31] for  $p \geq 3$  and Haagerup (1982) [11] for  $p \geq 2$ ; here and elsewhere, the  $\varepsilon_i$ 's are independent Rademacher random variables (r.v.'s), so that  $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = 1/2$  for all  $i$ , and  $Z \sim N(0, 1)$ .

For  $f(x) = e^{\lambda x}$  ( $\lambda \geq 0$ ), this inequality follows from Hoeffding (1963) [12], whence

$$P(\varepsilon_1 a_1 + \dots + \varepsilon_n a_n \geq x) \leq \inf_{\lambda \geq 0} \frac{E e^{\lambda Z}}{e^{\lambda x}} = e^{-x^2/2}, \quad x \geq 0.$$

Since  $P(Z \geq x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$  ( $x \rightarrow \infty$ ), a factor  $\asymp \frac{1}{x}$  is "missing" here. The apparent cause of this deficiency is that the class of the exponential moment functions  $f(x) = e^{\lambda x}$  ( $\lambda \geq 0$ ) is too small (and so is the class of the power functions  $f(x) = |x|^p$ ).

Consider the much richer classes of functions  $\mathcal{F}_+^{(\alpha)}$  ( $\alpha \geq 0$ ), consisting of all the functions  $f: \mathbb{R} \rightarrow \mathbb{R}$  given by the formula

$$f(x) = \int_{-\infty}^{\infty} (x - t)_+^{\alpha} \mu(dt), \quad u \in \mathbb{R},$$

where  $\mu \geq 0$  is a Borel measure,  $x_+ := \max(0, x)$ ,  $x_+^{\alpha} := (x_+)^{\alpha}$ ,  $0^0 := 0$ .

It is easy to see [25, Proposition 1(ii)] that

$$(1.2) \quad 0 \leq \beta < \alpha \quad \text{implies} \quad \mathcal{F}_+^{(\alpha)} \subseteq \mathcal{F}_+^{(\beta)}.$$

**Proposition 1.1.** [28] *For natural  $\alpha$ , one has  $f \in \mathcal{F}_+^{(\alpha)}$  if and only if  $f$  has finite derivatives  $f^{(0)} := f, f^{(1)} := f', \dots, f^{(\alpha-1)}$  on  $\mathbb{R}$  such that  $f^{(j)}(-\infty) = 0$  for  $j = 0, 1, \dots, \alpha - 1$  and  $f^{(\alpha-1)}$  is convex on  $\mathbb{R}$ .*

It follows from Proposition 1.1 that, for every  $t \in \mathbb{R}$ , every  $\beta \geq \alpha$ , and every  $\lambda > 0$ , the functions  $u \mapsto (u - t)_+^{\beta}$  and  $u \mapsto e^{\lambda(u-t)}$  belong to  $\mathcal{F}_+^{(\alpha)}$ , while the functions  $u \mapsto |u - t|^{\beta}$  and  $u \mapsto \cosh \lambda(u - t)$  belong to  $\mathcal{F}^{(\alpha)}$ .

Eaton (1970) [6] proved the Khinchin-Whittle-Haagerup inequality (1.1) for a class of moment functions, which essentially coincides with the class  $\mathcal{F}_+^{(3)}$ . Based on asymptotics, numerics, and a certain related inequality, Eaton (1974) [7] conjectured that the mentioned moment comparison inequality of his implies that

$$P(\varepsilon_1 a_1 + \dots + \varepsilon_n a_n \geq x) \leq \frac{2e^3}{9} \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} \quad \forall x > \sqrt{2}.$$

Pinelis (1994) [23] proved the following improvement of this conjecture:

$$(1.3) \quad P(\varepsilon_1 a_1 + \dots + \varepsilon_n a_n \geq x) \leq \frac{2e^3}{9} P(Z \geq x) \quad \forall x \in \mathbb{R},$$

as well as certain multidimensional extensions of these results.

Later it was realized in Pinelis (1998) [24] that the reason why it is possible to extract tail comparison inequality (1.3) from the Khinchin-Eaton moment comparison inequality (1.1) for  $f \in \mathcal{F}_+^{(3)}$  is that the tail function  $x \mapsto \mathbf{P}(Z \geq x)$  is log-concave. This realization resulted in a general device, which allows one to extract the optimal tail comparison inequality from an appropriate moment comparison inequality. The following is a special case of Theorem 4 of Pinelis (1999) [25]; see also Theorem 3.11 of Pinelis (1998) [24].

**Theorem 1.2.** *Suppose that  $0 \leq \beta \leq \alpha$ ,  $\xi$  and  $\eta$  are real-valued r.v.'s, and the tail function  $u \mapsto \mathbf{P}(\eta \geq u)$  is log-concave on  $\mathbb{R}$ . Then the comparison inequality*

$$(1.4) \quad \mathbf{E}f(\xi) \leq \mathbf{E}f(\eta) \quad \text{for all } f \in \mathcal{F}_+^{(\alpha)}$$

*implies*

$$(1.5) \quad \mathbf{E}f(\xi) \leq c_{\alpha,\beta} \mathbf{E}f(\eta) \quad \text{for all } f \in \mathcal{F}_+^{(\beta)}$$

*and, in particular, for all real  $x$ ,*

$$(1.6) \quad \mathbf{P}(\xi \geq x) \leq \inf_{f \in \mathcal{F}_+^{(\alpha)}} \frac{\mathbf{E}f(\eta)}{f(x)}$$

$$(1.7) \quad = B_{\text{opt}}(x) := \inf_{t \in (-\infty, x)} \frac{\mathbf{E}(\eta - t)_+^\alpha}{(x - t)^\alpha}$$

$$(1.8) \quad \leq \min \left( c_{\alpha,0} \mathbf{P}(\eta \geq x), \inf_{h>0} e^{-hx} \mathbf{E}e^{h\eta} \right),$$

*where*

$$(1.9) \quad c_{\alpha,\beta} := \frac{\Gamma(\alpha+1)(e/\alpha)^\alpha}{\Gamma(\beta+1)(e/\beta)^\beta}.$$

*Moreover, the constant  $c_{\alpha,\beta}$  is the best possible in (1.5) and (1.9).*

A similar result for the case when  $\alpha = 1$  and  $\beta = 0$  is contained in the book by Shorack and Wellner (1986) [30], pages 797–799.

*Remark 1.3.* As follows from [24, Remark 3.13], a useful point is that the requirement of the log-concavity of the tail function  $q(u) := \mathbf{P}(\eta \geq u)$  in Theorem 1.2 can be relaxed by replacing  $q(x) = \mathbf{P}(\eta \geq x)$  by any [e.g., the least] log-concave majorant of  $q$ . However, then the optimality of  $c(\alpha, \beta)$  is then not guaranteed.

Note that  $c_{3,0} = 2e^3/9$ , which is the constant factor in (1.3). Bobkov, Götze, and Houdré (2001) [5] obtained a simpler proof of inequality (1.3), but with a constant factor  $12.0099\dots$  in place of  $2e^3/9 = 4.4634\dots$

Pinelis (1999) [25] obtained the “discrete” improvement of (1.3):

$$(1.10) \quad \mathbf{P}(\varepsilon_1 a_1 + \dots + \varepsilon_n a_n \geq x) \leq \frac{2e^3}{9} \mathbf{P}\left(\frac{1}{\sqrt{n}}(\varepsilon_1 + \dots + \varepsilon_n) \geq x\right)$$

for all values  $x$  of r.v.  $\frac{1}{\sqrt{n}}(\varepsilon_1 + \dots + \varepsilon_n)$ .

## 2. DOMINATION BY NORMAL MOMENTS AND TAILS

**Theorem 2.1.** *Let  $S_0 \leq 0, S_1, \dots$  be a supermartingale, with increments  $X_i := S_i - S_{i-1}$ ,  $i = 1, 2, \dots$ . Suppose that for every  $i = 1, 2, \dots$  there exist  $H_{\leq(i-1)}$ -measurable r.v.'s  $C_{i-1}$  and  $D_{i-1}$  and a positive real number  $s_i$  such that*

$$(2.1) \quad C_{i-1} \leq X_i \leq D_{i-1} \quad \text{and}$$

$$(2.2) \quad D_{i-1} - C_{i-1} \leq 2s_i$$

*with probability 1. Then for all  $f \in \mathcal{F}_+^{(5)}$  and all  $n = 1, 2, \dots$*

$$(2.3) \quad \mathbb{E}f(S_n) \leq \mathbb{E}f(sZ),$$

*where*

$$s := \sqrt{s_1^2 + \dots + s_n^2}$$

*and  $Z \sim N(0, 1)$ .*

The proof of this and other statements (whenever necessary) are deferred to Section 5.

By virtue of Theorem 1.2, one has the following corollary under the conditions of Theorem 2.1.

**Corollary 2.2.** *For all  $\beta \in [0, 5]$ , all  $f \in \mathcal{F}_+^{(\beta)}$ , and all  $n = 0, 1, \dots$*

$$(2.4) \quad \mathbb{E}f(S_n) \leq c_{5,\beta} \mathbb{E}f(sZ).$$

*In particular, for all real  $x$ ,*

$$(2.5) \quad \mathbb{P}(S_n \geq x) \leq \inf_{f \in \mathcal{F}_+^{(5)}} \frac{\mathbb{E}f(sZ)}{f(x)}$$

$$(2.6) \quad = \inf_{t \in (-\infty, x)} \frac{\mathbb{E}(sZ - t)_+^\alpha}{(x - t)^\alpha}$$

$$(2.7) \quad \leq \min \left( c_{5,0} \mathbb{P}(sZ \geq x), \inf_{h>0} e^{-hx} \mathbb{E}e^{hsZ} \right)$$

$$(2.8) \quad = \min \left( c_{5,0} \overline{\Phi} \left( \frac{x}{s} \right), \exp \left( -\frac{x^2}{2s^2} \right) \right),$$

*and*

$$c_{5,0} = 5!(e/5)^5 = 5.699 \dots$$

The upper bound  $\exp \left( -\frac{x^2}{2s^2} \right)$  was obtained by Hoeffding (1963) [12] for the case when the  $C_{i-1}$ 's and  $D_{i-1}$ 's are non-random.

The upper bound (2.7) – but with constant factor 435 in place of  $c_{5,0} = 5.699 \dots$  – was obtained in [1] for the case when  $(S_i)$  is a martingale.

**Theorem 2.3.** *Let  $S_0 \leq 0, S_1, \dots$  be a supermartingale, with increments  $X_i := S_i - S_{i-1}$ ,  $i = 1, 2, \dots$ . Suppose that for every  $i = 1, 2, \dots$  there exist*

a positive  $H_{\leq(i-1)}$ -measurable r.v.  $D_{i-1}$  and a positive real number  $s_i$  such that

$$(2.9) \quad X_i \leq D_{i-1} \quad \text{and}$$

$$(2.10) \quad \frac{1}{2} \left( D_{i-1} + \frac{\text{Var}_{i-1} X_i}{D_{i-1}} \right) \leq \hat{s}_i$$

with probability 1. Let

$$(2.11) \quad \hat{s} := \sqrt{\hat{s}_1^2 + \cdots + \hat{s}_n^2}.$$

Then one has all the inequalities (2.3)–(2.8), only with  $s$  replaced by  $\hat{s}$ .

*Remark 2.4.* Theorem 2.1 may be considered as a special case Theorem 2.3. Indeed, it can be seen from the proofs of these two theorems (see Lemma 5.1.1 and Lemma 3.1 in [28]), one may assume without loss of generality that the supermartingales  $(S_i)$  in Theorem 2.1 and 2.3 are actually martingales with  $S_0 = 0$ . Therefore, to deduce Theorem 2.1 from Theorem 2.3, it is enough to observe that for any r.v.  $X$  and constants  $c < 0$  and  $d > 0$ , one has the following implication:

$$(2.12) \quad EX = 0 \ \& \ \mathbf{P}(c \leq X \leq d) = 1 \implies \text{Var}X \leq |c|d.$$

In turn, implication (2.12) follows from [16], which reduces the situation to that of a r.v.  $X$  taking on only two values. Alternatively, in light of the duality result [24, (4)], it is easy to give a direct proof of (2.12). Indeed,  $EX = 0$  and  $\mathbf{P}(c \leq X \leq d) = 1$  imply

$$0 \geq \mathbf{E}(X - c)(X - d) = EX^2 + cd = \text{Var}X - |c|d.$$

However, rather than deducing Theorem 2.1 from Theorem 2.3, we shall go in the opposite direction, proving Theorem 2.3 based on Theorem 2.1.

Thus, Theorem 2.1 is seen as the main result of this paper.

*Remark 2.5.* The set of conditions (2.9)–(2.10) is equivalent to

$$X_i \leq D_{i-1} \quad \text{and} \quad \sigma_*(D_{i-1}, \mathbf{E}_{i-1} X_i^2) \leq s_i$$

with probability 1, where

$$\begin{aligned} \sigma_*(d_0, \sigma^2) &:= \frac{1}{2} \inf_{d \geq d_0} \left( d + \frac{\sigma^2}{d} \right) = \min \left( \sigma \vee d_0, \frac{1}{2} \left( d_0 + \frac{\sigma^2}{d_0} \right) \right) \\ &= \begin{cases} \sigma & \text{if } \sigma \geq d_0, \\ \frac{1}{2} \left( d_0 + \frac{\sigma^2}{d_0} \right) & \text{if } \sigma < d_0, \end{cases} \end{aligned}$$

for positive  $\sigma$  and  $d_0$ . This follows simply because the inequalities  $X_i \leq D_{i-1}$  and  $d \geq D_{i-1}$  imply  $X_i \leq d$ .

From the “right-tail” bounds stated above, “two-tail” ones immediately follow:

**Corollary 2.6.** *Let  $S_0 = 0, S_1, \dots$  be a martingale, with increments  $X_i := S_i - S_{i-1}$ ,  $i = 1, 2, \dots$ . Suppose that conditions (2.1) and (2.2) hold. Then inequalities (2.3) and (2.4) hold for all  $f \in \mathcal{F}^{(5)}$  and  $f \in \mathcal{F}^{(\beta)}$  ( $\beta \in [0, 5]$ ), rather than only for all  $f \in \mathcal{F}_+^{(5)}$  and  $f \in \mathcal{F}_+^{(\beta)}$ , respectively.*

**Corollary 2.7.** *Let  $S_0 = 0, S_1, \dots$  be a martingale, with increments  $X_i := S_i - S_{i-1}$ ,  $i = 1, 2, \dots$ . Suppose that condition (2.10) holds, and condition (2.9) holds for  $|X_i|$  in place of  $X_i$ . Then inequalities (2.3) and (2.4) with  $s$  replaced by  $\hat{s}$  hold for all  $f \in \mathcal{F}^{(5)}$  and  $f \in \mathcal{F}^{(\beta)}$  ( $\beta \in [0, 5]$ ), rather than only for all  $f \in \mathcal{F}_+^{(5)}$  and  $f \in \mathcal{F}_+^{(\beta)}$ , respectively.*

That  $(S_0, S_1, \dots)$  in Theorems 2.1 and 2.3 is allowed to be a supermartingale (rather than only a martingale) makes it convenient to use the simple but powerful truncation tool. (Such a tool was used, for example, in [22] to prove limit theorems for large deviation probabilities based only on precise enough probability inequalities and without using Cramér's transform, the standard device in the theory of large deviations.) Thus, for instance, one has the following corollary from Theorem 2.3.

**Corollary 2.8.** *Let  $S_0 \leq 0, S_1, \dots$  be a supermartingale, with increments  $X_i := S_i - S_{i-1}$ ,  $i = 1, 2, \dots$ . For every  $i = 1, 2, \dots$ , let  $D_{i-1}$  be a positive  $\mathcal{H}_{\leq(i-1)}$ -measurable r.v. and let  $s_i$  be a positive real number such that (2.10) holds (while (2.9) does not have to). Let  $\hat{s}$  be still defined by (2.11).*

*Then for all real  $x$*

(2.13)

$$\begin{aligned} \mathbb{P}(S_n \geq x) &\leq \mathbb{P}\left(\max_{1 \leq i \leq n} \frac{X_i}{D_{i-1}} \geq 1\right) + \min\left(c_{5,0} \overline{\Phi}\left(\frac{x}{\hat{s}}\right), \exp\left(-\frac{x^2}{2\hat{s}^2}\right)\right) \\ (2.14) \quad &\leq \sum_{1 \leq i \leq n} \mathbb{P}(X_i \geq D_{i-1}) + \min\left(c_{5,0} \overline{\Phi}\left(\frac{x}{\hat{s}}\right), \exp\left(-\frac{x^2}{2\hat{s}^2}\right)\right). \end{aligned}$$

These bounds are much more precise than the exponential bounds in [10, 9, 20].

### 3. MAXIMAL INEQUALITIES

Introduce

$$M_n := \max_{0 \leq k \leq n} S_k.$$

**Theorem 3.1.** *Let  $(S_0 = 0, S_1, \dots)$  be a martingale. Then the upper bounds on  $\mathbb{P}(S_n \geq x)$  given in Corollary 2.2 and Theorem 2.3 are also upper bounds on  $\mathbb{P}(M_n \geq x)$ , under the same conditions: (2.1)-(2.2) and (2.9)-(2.10), respectively.*

**Theorem 3.2.** *Let  $0 \leq \beta \leq \alpha$  and  $x > t$ , and let  $(S_n)$  be a martingale or, more generally, a submartingale. Assume, moreover, that  $\alpha > 1$ . Then, for*

any natural  $n$ ,

$$(3.1) \quad \mathbb{E}(M_n - x)_+^\beta \leq k_{1;\alpha,\beta} \frac{\mathbb{E}(S_n - t)_+^\alpha}{(x - t)^{\alpha-\beta}},$$

where

$$(3.2) \quad k_{1;\alpha,\beta} := \sup_{\sigma > 0} \sigma^{-\beta(\alpha-1)} \left( \int_0^\sigma \frac{\beta s^{\beta-1} ds}{1+s} \right)^\alpha$$

if  $\beta > 0$ , and  $k_1(\alpha, 0) := 1$ . The particular cases of (3.1), corresponding to  $\beta = 0$  and  $\beta = \alpha$ , respectively, are Doob's inequalities

$$(3.3) \quad \mathbb{P}(M_n \geq x) \leq \frac{\mathbb{E}(S_n - t)_+^\alpha}{(x - t)^\alpha}$$

and

$$(3.4) \quad \mathbb{E}(M_n)_+^\alpha \leq \left( \frac{\alpha}{\alpha-1} \right)^\alpha \mathbb{E}(S_n)_+^\alpha.$$

**Theorem 3.3.** *Let  $(S_0 = 0, S_1, \dots)$  be a martingale. Then inequalities (2.4) and (2.5) hold if  $S_n$  is replaced there by  $M_n$  and  $c_{5,\beta}$  by  $\frac{k_{1;\alpha,\beta}}{k_{\alpha,\beta}} c_{5,\beta}$ , under the same conditions: (2.1)-(2.2) and (2.9)-(2.9), respectively.*

Similarly, results of [28] can be extended.

*Remark 3.4.* Note that

$$\int_0^\sigma \frac{\beta s^{\beta-1} ds}{1+s} = \sigma^\beta {}_2F_1(\beta, 1; 1+\beta; -\sigma) = \beta \int_0^1 (1-u)^{\beta-1} (1+\sigma u)^{-\beta} du,$$

where  ${}_2F_1$  is a hypergeometric function. Note also that there is some  $\sigma_{\alpha,\beta} \in (0, \infty)$  such that the expression under the sup sign in (3.2) is increasing in  $\sigma \in (0, \sigma_{\alpha,\beta})$  and decreasing in  $\sigma \in (\sigma_{\alpha,\beta}, \infty)$ ; this can be seen from the proof of Proposition 3.9. Thus, the sup is attained at the unique point  $\sigma_{\alpha,\beta}$ .

**Proposition 3.5.** *Let  $\alpha$  and  $\beta$  be as in Theorem 3.2. Then*

$$(3.5) \quad k_{1;\alpha,\beta} \leq k_{2;\alpha,\beta} := \frac{\Gamma(1+\beta)\Gamma(\alpha-\beta)}{\Gamma(\alpha)}.$$

*Remark 3.6.*

$$k_2(\alpha, 0) = k(\alpha, 0) = 1 = k_1(\alpha, 0).$$

**Proposition 3.7.** *Let  $0 \leq \beta < \alpha$ ,  $x > t$ , and*

$$(3.6) \quad k_{\alpha,\beta} := \frac{\beta^\beta (\alpha - \beta)^{\alpha-\beta}}{\alpha^\alpha}.$$

*Then*

$$(3.7) \quad \forall u \in \mathbb{R} \quad (u - x)_+^\beta \leq k_{\alpha,\beta} \frac{(u - t)_+^\alpha}{(x - t)^{\alpha-\beta}},$$

*and  $k_{\alpha,\beta}$  is the best constant here. (The values at  $\beta = 0$  are understood here as the corresponding limits as  $\beta \downarrow 0$ .)*

**Proposition 3.8.** *Let  $0 \leq \beta \leq \alpha$  and  $x > t$ , and let  $(S_n)$  be a martingale or, more generally, a submartingale. Then, for any natural  $n$ ,*

$$(3.8) \quad \mathbb{E}(S_n - x)_+^\beta \leq k_{\alpha,\beta} \frac{\mathbb{E}(S_n - t)_+^\alpha}{(x - t)^{\alpha-\beta}},$$

and  $k_{\alpha,\beta}$  is the best constant here.

**Proposition 3.9.** *Let  $\alpha$  and  $\beta$  be as in Theorem 3.2. Then*

$$(3.9) \quad k_{1;\alpha,\beta} \leq k_{3;\alpha,\beta} := k_{\alpha,\beta} \left( \frac{\alpha}{\alpha - 1} \right)^\alpha,$$

where  $k_{\alpha,\beta}$  is defined by (3.6).

**Proposition 3.10.** *Let  $\alpha > 1$ . Then*

$$(3.10) \quad k_1(\alpha, \alpha) = k_3(\alpha, \alpha) = \left( \frac{\alpha}{\alpha - 1} \right)^\alpha.$$

**Corollary 3.11.** *Let  $\alpha$  and  $\beta$  be as in Theorem 3.2. Then*

$$(3.11) \quad k(\alpha, \alpha) \leq k_1(\alpha, \alpha) \leq k_2(\alpha, \alpha) \wedge k_3(\alpha, \alpha);$$

at that

$$(3.12) \quad k(\alpha, 0) = k_1(\alpha, 0) = k_2(\alpha, 0) = 1,$$

while

$$(3.13) \quad k_1(\alpha, \alpha) = k_3(\alpha, \alpha) = \left( \frac{\alpha}{\alpha - 1} \right)^\alpha > k(\alpha, \alpha) = 1.$$

#### 4. CONCENTRATION INEQUALITIES FOR SEPARATELY LIPSCHITZ FUNCTIONS

**Definition 4.1.** Let us say that a real-valued function  $g$  of  $n$  (not necessarily real-valued) arguments is *separately Lipschitz* if it satisfies a Lipschitz type condition in each of its arguments:

$$(4.1) \quad |g(x_1, \dots, x_{i-1}, \tilde{x}_i, x_{i+1}, \dots, x_n) - g(x_1, \dots, x_n)| \leq \rho_i(\tilde{x}_i, x_i) < \infty$$

for all  $i$  and all  $x_1, \dots, x_n, \tilde{x}_i$ , where  $\rho_i(\tilde{x}_i, x_i)$  depends only on  $\tilde{x}_i$  and  $x_i$ . Let the *radius* of the separately Lipschitz function  $g$  be defined as

$$r := \sqrt{r_1^2 + \dots + r_n^2},$$

where

$$(4.2) \quad r_i := \frac{1}{2} \sup_{\tilde{x}_i, x_i} \rho_i(\tilde{x}_i, x_i).$$

The concentration inequalities given in this section follow from martingale inequalities given in Section 2. Their proofs here are based on the improvements given in [21] and [29] of the method of Yurinskii (1974) [32]; cf. [18, 19] and [1].



Papers [32], [21], and [29] deal mainly with separately Lipschitz function  $g$  of the form

$$g(x_1, \dots, x_n) = \|x_1 + \dots + x_n\|,$$

where the  $x_i$ 's are vectors in a normed space; however, it was already understood there that the methods would work for much more general functions  $g$  – see [29, Remark 1]. In a similar fashion, various concentration inequalities for general functions  $g$  were obtained in [18, 19] and [1].

**Theorem 4.2.** *Suppose that a r.v.  $Y$  can be represented as a real-valued function  $g$  of independent (not necessarily real-valued) r.v.'s  $X_1, \dots, X_n$ :*

$$Y = g(X_1, \dots, X_n),$$

where  $g$  is separately Lipschitz with radius  $r$ . Then

$$(4.3) \quad \mathbb{E}f(Y - \mathbb{E}Y) \leq \mathbb{E}f(rZ) \quad \text{for all } f \in \mathcal{F}^{(5)} \quad \text{and}$$

$$(4.4) \quad \mathbb{E}f(Y - \mathbb{E}Y) \leq c_{5,\beta} \mathbb{E}f(rZ) \quad \text{for all } \beta \in [0, 5] \quad \text{and all } f \in \mathcal{F}^{(\beta)},$$

where  $Z \sim N(0, 1)$ . In particular, for all real  $x$ ,

$$(4.5) \quad \mathbb{P}(Y - \mathbb{E}Y \geq x) \leq c_{5,0} \mathbb{P}(rZ \geq x) = c_{5,0} \overline{\Phi}\left(\frac{x}{r}\right).$$

**Proposition 4.3.** *Inequalities (4.3), (4.4), and (4.5) will hold if the conditions of Theorem 4.2 are relaxed so that  $r_i$  is replaced by*

$$(4.6) \quad \hat{r}_i := \frac{1}{2} \sup_{x_1, \dots, x_i, \tilde{x}_i} |\mathbb{E}g(x_1, \dots, x_{i-1}, \tilde{x}_i, X_{i+1}, \dots, X_n) - \mathbb{E}g(x_1, \dots, x_i, X_{i+1}, \dots, X_n)|,$$

for every  $i$ . Note that  $\hat{r}_i \leq r_i$  for all  $i$ .

*Remark 4.4.* The upper bound given by (4.5) can be replaced by the tighter bound

$$\min\left(\exp\left(-\frac{x^2}{2r^2}\right), c_{5,0} \overline{\Phi}\left(\frac{x}{r}\right)\right),$$

which is less than  $\exp\left(-\frac{x^2}{2r^2}\right)$  for all  $\frac{x}{r} \geq 1.89$ .

The foregoing conditions can be modified as follows.

**Theorem 4.5.** *Suppose that*

$$(4.7) \quad \Xi_i(x_1, \dots, x_{i-1}, x_i) := \mathbb{E}g(x_1, \dots, x_{i-1}, x_i, X_{i+1}, \dots, X_n) - \mathbb{E}g(x_1, \dots, x_{i-1}, X_i, X_{i+1}, \dots, X_n)$$

$$(4.8) \quad \leq D_{i-1}(x_1, \dots, x_{i-1}),$$

and

$$(4.9) \quad \frac{1}{2} \left( D_{i-1}(x_1, \dots, x_{i-1}) + \frac{\mathbb{E}\Xi_i(x_1, \dots, x_{i-1}, X_i)^2}{D_{i-1}(x_1, \dots, x_{i-1})} \right) \leq s_i,$$

for all  $i$  and all  $x_1, \dots, x_{i-1}, x_i$ , where  $D_{i-1} > 0$  depends only on  $i$  and  $x_1, \dots, x_{i-1}$ , and  $s_i$  depends only on  $i$ . Let

$$s := \sqrt{s_1^2 + \dots + s_n^2}.$$

Then inequalities (4.3), (4.4), and (4.5) will hold if  $r$  is replaced there by  $s$ .

The next two propositions show how to obtain good upper bounds on  $\Xi_i(x_1, \dots, x_{i-1}, x_i)$  and  $\mathbb{E}\Xi_i(x_1, \dots, x_{i-1}, X_i)^2$ , to be used in Theorem 4.5.

**Proposition 4.6.** *If  $g$  is separately Lipschitz so that (4.1) holds, then for all  $i$  and all  $x_1, \dots, x_{i-1}$ ,*

$$(4.10) \quad \mathbb{E}\Xi_i(x_1, \dots, x_{i-1}, X_i)^2 \leq \inf_{x_i} \mathbb{E}\rho_i(X_i, x_i)^2 \leq \mathbb{E}\rho_i(X_i, \mathbb{E}X_i)^2.$$

*If, moreover, the function  $g$  is convex in each of its arguments, then for all  $i$  and all  $x_1, \dots, x_i$ ,*

$$(4.11) \quad \Xi_i(x_1, \dots, x_{i-1}, x_i) \leq \rho_i(x_i, \mathbb{E}X_i).$$

*Remark 4.7.* We do not require that  $\rho_i$  be a metric. However, the smallest possible  $\rho_i$ , which is the supremum of the left-hand side of (4.1) over all  $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ , is necessarily a metric. Note also that, for  $r_i$  defined by (4.1),

$$\rho_i(x_i, \mathbb{E}X_i) = \rho_i(x_i, 0) \leq \frac{1}{2} r_i$$

for all  $x_i$ , provided the following conditions: (i)  $\rho_i$  is the smallest possible and, moreover, is a norm; (ii)  $X_i$  is symmetrically distributed; and (iii)  $x_i$  belongs to the support of the distribution of  $X_i$ .

**Corollary 4.8.** *Let here  $X_1, \dots, X_n$  be independent r.v.'s with values in a separable Banach space with norm  $\|\cdot\|$ , and let*

$$Y := \|X_1 + \dots + X_n\|.$$

*Suppose that, with probability 1,*

$$(4.12) \quad \|X_i - \mathbb{E}X_i\| \leq d_i$$

*and*

$$(4.13) \quad \frac{1}{2} \left( d_i + \frac{\mathbb{E}\|X_i - \mathbb{E}X_i\|^2}{d_i} \right) \leq s_i,$$

*for all  $i$ , where  $d_i > 0$  and  $s_i > 0$  are non-random constants. Let*

$$s := \sqrt{s_1^2 + \dots + s_n^2}.$$

*Then inequalities (4.3), (4.4), and (4.5) will hold if  $r$  is replaced there by  $s$ .*

## 5. PROOFS

**5.1. Proofs for Section 2.** Let us first observe that Theorem 2.1 can be easily reduced to the case when  $(S_n)$  is a martingale. This is implied by the following two lemmas.

The next lemma is obvious and stated here for the convenience of reference.

**Lemma 5.1.1.** *Let  $(S_n)$  be a supermartingale as in Theorem 2.1, so that conditions (2.1) and (2.2) are satisfied. Let*

$$\tilde{X}_i := X_i - \mathbf{E}_{i-1}X_i, \quad \tilde{C}_{i-1} := C_{i-1} - \mathbf{E}_{i-1}X_i, \quad \text{and} \quad \tilde{D}_{i-1} := D_{i-1} - \mathbf{E}_{i-1}X_i.$$

*Then  $\tilde{X}_i$  is  $H_{\leq i}$ -measurable,  $\tilde{C}_{i-1}$  and  $\tilde{D}_{i-1}$  are  $H_{\leq(i-1)}$ -measurable, and one has*

$$\begin{aligned} X_i &\leq \tilde{X}_i, \\ \mathbf{E}_{i-1}\tilde{X}_i &= 0, \\ \tilde{C}_{i-1} &\leq \tilde{X}_i \leq \tilde{D}_{i-1}, \quad \text{and} \\ \tilde{D}_{i-1} - \tilde{C}_{i-1} &\leq 2s_i \end{aligned}$$

*with probability 1.*

*Proof of Theorem 2.1.* is similar to the proof of Theorem 2.1 in [28] but based on the following lemma, in place of Lemma 3.2 in [28]. (Also, one has to refer here to Lemma 5.1.1 instead of Lemma 3.1 in [28].)  $\square$

**Lemma 5.1.2.** *Let  $X$  be a r.v. such that  $\mathbf{E}X = 0$  and  $c \leq X \leq d$  with probability 1 for some real constants  $c$  and  $d$  (whence  $c \leq 0$  and  $d \geq 0$ ). Let  $Z \sim N(0, 1)$ . Then for all  $f \in \mathcal{F}_+^{(5)}$*

$$(5.1) \quad \mathbf{E}f(X) \leq \mathbf{E}f((d - c)Z).$$

*Proof.* This proof is rather long. Let  $\mathcal{X}_{c,d}$  be the set of all r.v.'s  $X$  such that  $\mathbf{E}X = 0$  and  $c \leq X \leq d$  with probability 1. In view of [16] (say), for any given real  $t$ , a maximum of  $\mathbf{E}f_t(X)$  over all r.v.'s  $X$  in  $\mathcal{X}_{c,d}$  is attained when  $X$  takes on only two values, say  $a$  and  $b$ , in the interval  $[c, d]$ . Since the function  $f_t$  is convex, it then follows that, without loss of generality (w.l.o.g.),  $a = c$  and  $b = d$ . (Indeed,  $\mathbf{E}g(\sigma Z)$  is non-decreasing in  $\sigma > 0$  for  $Z \sim N(0, 1)$  and any convex function  $g$ . One way to verify the latter statement is as follows. It suffices to consider the functions of the form  $g(u) = (u - t)_+$  for real  $t$ ; cf. identity ( ) in Pinelis (1994). But the derivative of  $\mathbf{E}(\sigma Z - t)_+$  in  $\sigma > 0$  is  $\varphi(t/\sigma) > 0$ . Alternatively, one can prove that  $\mathbf{E}g(\sigma Z)$  is non-decreasing in  $\sigma > 0$  by an application of Jensen's inequality.) Moreover, by rescaling, w.l.o.g.  $d - c = 2$ . In other words, then one has the following:

$$X = \begin{cases} 2r & \text{with probability } 1 - r, \\ 2r - 2 & \text{with probability } r, \end{cases}$$

for some  $r \in [0, 1]$ . At that,

$$Y \sim N(0, 1).$$

Now the right-hand side of inequality (5.1) can be written as

$$(5.2) \quad \mathbb{E}f_t(Y) = R(t) := P(t)\varphi(t) - Q(t)\overline{\Phi}(t),$$

where

$$P(t) := 8 + 9t^2 + t^4 \quad \text{and} \quad Q(t) := t(15 + 10t^2 + t^4),$$

and its left-hand side as

$$(5.3) \quad \mathbb{E}f_t(X) = L(r, t) := r(2r - 2 - t)_+^5 + (1 - r)(2r - t)_+^5,$$

so that (5.1) is reduced to the inequality

$$(5.4) \quad L(r, t) \leq R(t)$$

for all  $r \in [0, 1]$  and all real  $t$ .

Note that (5.4) is trivial for  $t \geq 2r$ , because then  $L(r, t) = 0$ .

Therefore, it remains to consider two cases:  $(r, t) \in B$  and  $(r, t) \in C$ , where

$$B := \{(r, t) : 0 \leq r \leq 1, t \leq 2r - 2\} \quad \text{and}$$

$$C := \{(r, t) : 0 \leq r \leq 1, 2r - 2 \leq t \leq 2r\}.$$

**Case 1**  $(r, t) \in B$ . Note that in this case  $t \leq 0$  and, by (5.3),

$$L(r, t) = r(2r - 2 - t)^5 + (1 - r)(2r - t)^5.$$

For  $t \neq 0$ , one has the identity

$$(5.5) \quad \frac{Q(t)^2}{\varphi(t)} \partial_t \left( \frac{R(t) - L(r, t)}{Q(t)} \right) = Q_2(r, t) := \frac{Q_1(r, t)}{\varphi(t)} - 120,$$

where

$$Q_1(r, t) := Q'(t)L(r, t) - Q(t)\partial_t L(r, t),$$

which is a polynomial in  $r$  and  $t$ . Note that

$$\partial_r Q_2(r, t) = \frac{\partial_r Q_1(r, t)}{\varphi(t)} \quad \text{and} \quad \partial_t Q_2(r, t) = \frac{20 Q(t)}{\varphi(t)} d(r, t),$$

where

$$d(r, t) := \frac{tQ_1(t) + \partial_t Q_1(t)}{20 Q(t)}$$

is a polynomial in  $r$  and  $t$ , of degree 2 in  $r$ . Therefore, the critical points of  $Q_2$  in the interior  $\text{int } B$  of domain  $B$  are the solutions  $(r, t)$  of the system of polynomial equations

$$\begin{cases} d(r, t) &= 0, \\ \partial_r Q_1(r, t) &= 0. \end{cases}$$

Further, one has

$$d(r, t) = 0 \text{ if and only if } r = r_1(u) \text{ or } r = r_2(u),$$

where

$$u := 2 - r - t > 0, \quad r_1(u) := \frac{1 + u/2}{1 + u} \in (0, 1), \quad \text{and} \quad r_2(u) := \frac{2 + 2u + u^2/2}{2 + 2u + u^2} \in (0, 1).$$

Using the Sturm theorem or the convenient command **Reduce** of Mathematica 5.0, one can see that the only solution  $u = u_1 > 0$  of the algebraic equation  $\partial_r Q_1(r, t)|_{r=r_1(u), t=2-r_1(u)-u} = 0$  is  $0.269\dots$ , and

$Q_2(r, t)|_{r=r_1(u_1), t=2-r_1(u_1)-u_1} < 0$ . As for the equation  $\partial_r Q_1(r, t)|_{r=r_2(u), t=2-r_2(u)-u} = 0$ , it has no solutions  $u > 0$ .

Thus,  $Q_2 < 0$  at the only critical point  $(r, t) = (r_1(u_1), 2 - r_1(u_1) - u_1)$  of  $Q_2$  in  $\text{int } B$ .

Next, with  $u > 0$ ,

$$Q_2(r, t)|_{r=0, t=2r-2-u} = -20 \left( 6 + \frac{(2+u)^5}{\varphi(2+u)} (7 + 4u + u^2) \right) < 0.$$

Similarly, with  $u > 0$ ,

$$Q_2(r, t)|_{r=1, t=2r-2-u} = -20 \left( 6 + \frac{u^5(3+u^2)}{\varphi(u)} \right) < 0.$$

Now consider the function

$$q_2(r) := Q_2(r, t)|_{t=2r-2}.$$

Then  $\varphi(2r-2)q_2'(r)$  is a polynomial, whose only root  $r = r_3 \in (0, 1)$  is  $0.865\dots$ . But  $q_2(r_3) < 0$ . Therefore,  $Q_2 < 0$  at the only critical point of  $Q_2$  in the relative interior of the boundary  $t = 2r - 2$  of domain  $B$ .

Thus, as far as the sign of  $Q_2$  on  $B$  is concerned, it remains to consider the behavior of  $Q_2$  as  $t \rightarrow -\infty$ , which is as follows:  $Q_2(r, t) \sim 20(2r-1)^2 t^7 \rightarrow -\infty < 0$  for every  $r \neq 1/2$  and  $Q_2(r, t) \sim 40t^3(5+t^2) \rightarrow -\infty < 0$  for  $r = 1/2$ .

(As usual,  $a \sim b$  means  $a/b \rightarrow 1$ .)

We conclude that  $Q_2 < 0$  on  $B$ . Hence, in view of (5.5), the ratio  $\frac{R(t)-L(r,t)}{Q(t)}$  is decreasing in  $t$  on  $B$ .

Next, note that  $\varphi(t)$  and  $1 - \overline{\Phi}(t)$  are  $o(1/|t|^p)$  for every  $p > 0$  as  $t \rightarrow -\infty$ . Hence, in view of (5.2), one has the following as  $t \rightarrow -\infty$ :  $R(t) - L(r, t) = -Q(t) - L(r, t) + o(1) \sim -10(2r-1)^2 t^3 \rightarrow \infty$  for every  $r \neq 1/2$  and  $R(t) - L(r, t) = -10t \rightarrow \infty$  for  $r = 1/2$ .

Hence,  $\frac{R(t)-L(r,t)}{Q(t)} < 0$  for each  $r \in (0, 1)$  and all  $t < 0$  with large enough  $|t|$ . Since  $\frac{R(t)-L(r,t)}{Q(t)}$  is decreasing in  $t$  on  $B$ , one has  $\frac{R(t)-L(r,t)}{Q(t)} < 0$  on  $B$ , whence  $L(r, t) \leq R(t)$  on  $B$  (because  $Q(t) \leq 0$  on  $B$ ).

It remains to consider

**Case 2**  $(r, t) \in C$ . Here, letting  $v := 2r - t$ , one has  $0 \leq v \leq 2$ , and, by (5.3),

$$L(r, t) = (1-r)(2r-t)^5.$$

Let us use here notation introduced in the above consideration of Case 1. Then

$$d(r, t)|_{t=2r-v} = -(1-r)v^3 \left(1 - \frac{r}{2}v\right) < 0$$

for  $(r, t) = (r, 2r - v) \in \text{int } C$ . This implies that  $Q_2$  has no critical points in  $\text{int } C$ .

Next, with  $v > 0$ ,

$$Q_2(r, t)|_{r=0, t=2r-v} = -20 \left(6 + \frac{v^5(3+v^2)}{\varphi(t)}\right) < 0.$$

On the boundaries  $r = 1$  and  $t = 2r$  of  $C$ , one has  $Q_2 = -120 < 0$ . The boundary  $t = 2r - 2$  of  $C$  is common with  $B$ , and it was shown above that  $Q_2 < 0$  on that boundary as well.

Thus,  $Q_2 < 0$  on  $C$ . Since  $Q(t) = 0$  only for  $t = 0$ , it follows that the ratio  $\frac{R(t)-L(r,t)}{Q(t)}$  is decreasing in  $t$  on  $C$ .

Hence, just as on  $B$ , one has that  $L(r, t) < R(t)$  on  $C_- := \{(r, t) \in C : t \leq 0\}$ .

Moreover,  $\frac{R(t)-L(r,t)}{Q(t)} = \frac{R(t)}{Q(t)} > 0$  for  $t = 2r$ , since  $Q > 0$  on  $C_+ := C \setminus C_- = \{(r, t) \in C : t > 0\}$ . Because  $\frac{R(t)-L(r,t)}{Q(t)}$  is decreasing in  $t$ , one has  $\frac{R(t)-L(r,t)}{Q(t)} > 0$  on  $C_+$  and hence  $L(r, t) < R(t)$  on  $C_+$ .  $\square$

*Proof of Theorem 2.3.* This proof is similar to the proof of Theorem 2.1 in [28] and Theorem 2.1, but based on the following lemma, instead of Lemma 3.2 in [28] or 5.1.2. (As in the proof of Theorem 2.1, here one has also to refer to Lemma 3.1 in [28], rather than Lemma 5.1.1.)  $\square$

**Lemma 5.1.3.** *Suppose that  $X$  is a r.v. such that  $\mathbf{E}X = 0$ ,  $X \leq d$  with probability 1, and  $\mathbf{E}X^2 \leq \sigma^2$ , for some positive constants  $d$  and  $\sigma$ . Let*

$$s := \frac{1}{2} \left( d + \frac{\sigma^2}{d} \right).$$

*Let  $Z \sim N(0, 1)$ . Then for all  $f \in \mathcal{F}^{(5)}$*

$$(5.6) \quad \mathbf{E}f(X) \leq \mathbf{E}f(sZ).$$

*Proof.* In view of (1.2), one has  $\mathcal{F}^{(5)} \subseteq \mathcal{F}^{(2)}$ . Therefore, by Lemma 3.2 in [28], one may assume without loss of generality that here  $X = d \cdot X_a$ , where  $a = \sigma^2/d^2$ . Now it is seen that Lemma 5.1.3 follows from Lemma 5.1.2.  $\square$

## 5.2. Proofs for Section 3.

*Proof of Theorem 3.1.* Lemma 5.1.1 and Lemma 3.1 in [28] reduce Theorem 3.1 to the case when  $(S_n)$  is a martingale, and then Theorem 3.1 follows by Doob's inequality (3.3).  $\square$

*Proof of Theorem 3.2.* For every  $y > t$ , by Doob's inequality,

$$\mathbf{P}(M_n \geq y) \leq \frac{\mathbf{E}(S_n - t)_+ I\{M_n \geq y\}}{y - t}.$$

Hence, letting

$$(5.7) \quad J(u) := \int_x^u \frac{\beta(y-x)^{\beta-1}}{y-t} dy \, I\{u > x\} \quad \text{and} \quad \alpha' := \frac{\alpha}{\alpha-1},$$

and using Fubini's theorem, one has

$$\begin{aligned} \mathbb{E}(M_n - x)_+^\beta &= \int_x^\infty \beta(y-x)^{\beta-1} \mathbb{P}(M_n \geq y) dy \\ &\leq \int_x^\infty \beta(y-x)^{\beta-1} \frac{\mathbb{E}(S_n - t)_+ I\{M_n \geq y\}}{y-t} dy \\ &= \mathbb{E} \int_x^\infty \beta(y-x)^{\beta-1} \frac{(S_n - t)_+ I\{M_n \geq y\}}{y-t} dy \\ &= \mathbb{E}(S_n - t)_+ J(M_n) \\ (5.8) \quad &\leq (\mathbb{E}(S_n - t)_+^\alpha)^{1/\alpha} \left( \mathbb{E} J(M_n)^{\alpha'} \right)^{1/\alpha'}, \end{aligned}$$

by Hölder's inequality.

Observe that for all real  $u$

$$(5.9) \quad J(u) \leq c^{1/\alpha} (u-x)_+^{\beta/\alpha'}, \quad \text{where} \quad c := \frac{k_{1;\alpha,\beta}}{(x-t)^{\alpha-\beta}}.$$

Indeed, introducing new variables  $\sigma := \frac{u-x}{x-t}$  and  $s := \frac{y-x}{x-t}$ , one can see that, for  $u > x$ ,

$$\begin{aligned} J(u) &= (x-t)^{\beta-1} \int_0^\sigma \frac{\beta s^{\beta-1} ds}{1+s} \quad \text{and} \\ c^{1/\alpha} (u-x)_+^{\beta/\alpha'} &= k_{1;\alpha,\beta}^{1/\alpha} \sigma^{\beta(1-1/\alpha)} (x-t)^{\beta-1}, \end{aligned}$$

so that (5.9) follows, in view of (3.2).

Now (5.8) and (5.9) imply (3.8).  $\square$

*Proof of Theorem 3.3.* This is similar to the proof Theorem 3.1, but relies on inequality (3.1) in place of Doob's inequality (3.3).  $\square$

*Proof of Proposition 3.5.* Introduce

$$\begin{aligned} f(\sigma, \alpha, \beta, \gamma) &:= \sigma^{-\beta(\alpha-\gamma)/\gamma} \left( \int_0^\sigma \frac{\beta s^{\beta-1} ds}{(1+s)^\gamma} \right)^{\alpha/\gamma}, \\ K(\alpha, \beta, \gamma) &:= \sup_{\sigma > 0} f(\sigma, \alpha, \beta, \gamma). \end{aligned}$$

Then  $\sigma^{-\beta/\alpha} f(\sigma, \alpha, \beta, \gamma)^{1/\alpha} = (\mathbb{E} Y^\gamma)^{1/\gamma}$ , where  $Y := \frac{1}{1+S}$  and  $S$  is a r.v. with density  $s \mapsto \sigma^{-\beta} \beta s^{\beta-1} I\{0 < s < \sigma\}$ . Hence,  $f(\sigma, \alpha, \beta, \gamma)$  is non-decreasing in  $\gamma$ , and then so is  $K(\alpha, \beta, \gamma)$ . Therefore,

$$k_{1;\alpha,\beta} = K(\alpha, \beta, 1) \leq K(\alpha, \beta, \alpha) = k_{2;\alpha,\beta}.$$

$\square$

*Proof of Proposition 3.9.* By (3.2),

$$(5.10) \quad k_{1;\alpha,\beta} = \sup_{\sigma>0} r(\sigma)^\alpha,$$

where

$$r(\sigma) := \frac{f(\sigma)}{g(\sigma)}, \quad f(\sigma) := \int_0^\sigma \frac{\beta s^{\beta-1} ds}{1+s}, \quad \text{and} \quad g(s) := \sigma^{\beta(1-1/\alpha)}.$$

Note that the monotonicity pattern of

$$(5.11) \quad r_1(\sigma) := \frac{f'(\sigma)}{g'(\sigma)} = \frac{\alpha}{\alpha-1} \frac{\sigma^{\beta/\alpha}}{1+\sigma}$$

on  $(0, \infty)$  is  $\nearrow \searrow$ ; that is, there exists some  $\sigma_1(\alpha, \beta) \in (0, \infty)$  such that  $r_1$   $\nearrow$  (is increasing) on  $(0, \sigma_1(\alpha, \beta))$  and  $r_1$   $\searrow$  (is decreasing) on  $(\sigma_1(\alpha, \beta), \infty)$ ; namely, here

$$(5.12) \quad \sigma_1(\alpha, \beta) = \frac{\beta}{\alpha - \beta}.$$

Also,  $gg' > 0$  on  $(0, \infty)$ . Hence, it follows from [26, Proposition 1.9] that  $r$  has one of these monotonicity patterns on  $(0, \infty)$ :  $\nearrow$  or  $\searrow$  or  $\nearrow \searrow$  or  $\searrow \nearrow$  or  $\searrow \nearrow \searrow$ . However,  $r(\sigma)$  is positive on  $(0, \infty)$  and converges to 0 when  $\sigma \downarrow 0$  as well as when  $\sigma \rightarrow \infty$ . This leaves only one possible pattern for  $r$ :  $\nearrow \searrow$ . Hence, there is some  $\sigma(\alpha, \beta) \in (0, \infty)$ , at which  $r$  attains its maximum on  $(0, \infty)$ ; moreover,  $r'(\sigma(\alpha, \beta)) = 0$ , which is equivalent to  $r(\sigma(\alpha, \beta)) = r_1(\sigma(\alpha, \beta))$ . Thus,

$$\begin{aligned} k_{1;\alpha,\beta} &= \sup_{\sigma>0} r(\sigma)^\alpha = r(\sigma(\alpha, \beta))^\alpha = r_1(\sigma(\alpha, \beta))^\alpha \leq \sup_{\sigma>0} r_1(\sigma)^\alpha \\ &= r_1(\sigma_1(\alpha, \beta))^\alpha = k_{3;\alpha,\beta}, \end{aligned}$$

in view of (5.11), (5.12), and (3.9).  $\square$

*Proof of Proposition 3.10.* In the case  $\beta = \alpha > 1$ , the function  $r_1$  given by (5.11) is increasing on  $(0, \infty)$  to  $r_1(\infty) = \frac{\alpha}{\alpha-1}$ . Hence, so does  $r$ , according to the mentioned [26, Proposition 1.9]. Now Proposition 3.10 follows in view of (5.10).  $\square$

*Proof of Proposition 3.7.* Elementary calculus; the optimal value of  $u$ , when inequality (3.7) turns into an equality, is

$$(5.13) \quad u_* := \frac{\alpha x - \beta t}{\alpha - \beta} > x.$$

$\square$

*Proof of Proposition 3.8.* Only that  $k_{\alpha,\beta}$  is the best constant factor needs to be proved. Without loss of generality,  $x > 0$ . Suppose that (3.8) holds with some constant  $\tilde{k}$  in place of  $k_{\alpha,\beta}$ ; then, by continuity, it holds for



the continuous-time martingale  $S_v := B_{v \wedge \tau}$  in place of  $S_n$ , where  $B(\cdot)$  is a standard Brownian motion,  $v \geq 0$ , and

$$\tau := \inf\{v \geq 0: B_v = u_* \text{ or } B_v = \tilde{t}\};$$

here,  $u_*$  is defined by (5.13) and  $\tilde{t} := (-1) \wedge t$ . Note that  $\mathbf{E}\tau = u_*|\tilde{t}|$  and  $p := \frac{|\tilde{t}|}{|\tilde{t}| + u_*} > 0$ . It follows that

$$p \cdot (u_* - x)^\beta = \mathbf{E}(S_\infty - x)_+^\beta \leq \tilde{k} \frac{\mathbf{E}(S_\infty - t)_+^\alpha}{(x - t)^{\alpha - \beta}} = \tilde{k} p \frac{(u_* - t)_+^\alpha}{(x - t)^{\alpha - \beta}}.$$

Because  $k_{\alpha, \beta}$  is the best constant in (3.7), it follows now that  $\tilde{k} \geq k_{\alpha, \beta}$ .  $\square$

**5.3. Proofs for Section 4.** The proofs here are based on the improvements given in [21] and [29] of the method of Yurinskii (1974) [32]; cf. [18, 19] and [1].

For a r.v.  $Y$  as in Theorem 4.2, consider the martingale expansion

$$Y - \mathbf{E}Y = \xi_1 + \cdots + \xi_n,$$

of  $Y - \mathbf{E}Y$  with the martingale-differences

$$(5.14) \quad \xi_i := \mathbf{E}_i Y - \mathbf{E}_{i-1} Y.$$

where  $\mathbf{E}_i$  denotes the conditional expectation given  $H_{\leq i} := (X_1, \dots, X_i)$ . For each  $i$  pick an arbitrary non-random  $x_i$ , and introduce the r.v.

$$(5.15) \quad \eta_i := Y - \tilde{Y}_i, \quad \text{where} \quad \tilde{Y}_i := g(X_1, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_n).$$

*Proof of Theorem 4.2 and Proposition 4.3.* Note that, for the function  $\Xi_i$  defined by (4.7), one has  $\Xi_i(X_1, \dots, X_i) = \xi_i$ , where  $\xi_i$  is defined by (5.14). It follows from (5.14) that

$$(5.16) \quad C_{2,i-1} \leq \xi_i \leq D_{2,i-1} \quad \text{and} \quad D_{2,i-1} - C_{2,i-1} \leq 2\hat{r}_i \leq 2r_i,$$

where  $r_i$  and  $\hat{r}_i$  are given by (4.2) and (4.6), and

$$C_{2,i-1} := \inf_{x_i} \mathbf{E}_{i-1}(-\eta_i) = \inf_{x_i} \mathbf{E}_{i-1} \tilde{Y}_i - \mathbf{E}_{i-1} Y \quad \text{and} \\ D_{2,i-1} := \sup_{x_i} \mathbf{E}_{i-1}(-\eta_i) = \sup_{x_i} \mathbf{E}_{i-1} \tilde{Y}_i - \mathbf{E}_{i-1} Y$$

are  $H_{\leq(i-1)}$ -measurable. Now Proposition 4.3 – and hence Theorem 4.2 – follow by Theorem 2.1 and Corollary 2.2.  $\square$

*Proof of Theorem 4.5.* This proof is similar to that of Theorem 4.2 and Proposition 4.3, but based on Theorem 2.3 in place of Theorem 2.1 and Corollary 2.2. (Note that  $\mathbf{E}\Xi_i(x_1, \dots, x_{i-1}, X_i)^2$  is the same as conditional expectation  $\mathbf{E}_{i-1}\xi_i^2$  given that  $X_1 = x_1, \dots, X_{i-1} = x_{i-1}$ .)  $\square$

*Proof of Proposition 4.6.* For each  $i$ ,

$$(5.17) \quad \xi_i = \mathbf{E}_i \eta_i - \mathbf{E}_{i-1} \eta_i,$$

because  $\mathbb{E}_i \tilde{Y}_i = \mathbb{E}_{i-1} \tilde{Y}_i$ , in view of the independence of the  $X_i$ 's. Hence and by (4.1), for any given  $x_i$ ,

$$(5.18) \quad |\eta_i| \leq \rho_i(X_i, x_i)$$

with probability 1. It follows from (5.17) and (5.18) that, for any  $x_i$ ,

$$\begin{aligned} \mathbb{E}_{i-1} \xi_i^2 &= \mathbb{E}_{i-1} (\mathbb{E}_i \eta_i - \mathbb{E}_{i-1} \eta_i)^2 = \text{Var}_{i-1}(\mathbb{E}_i \eta_i) \leq \mathbb{E}_{i-1} (\mathbb{E}_i \eta_i)^2 \leq \mathbb{E}_{i-1} \mathbb{E}_i \eta_i^2 \\ &= \mathbb{E}_{i-1} \eta_i^2 \leq \mathbb{E}_{i-1} \rho_i(X_i, x_i)^2 = \mathbb{E} \rho_i(X_i, x_i)^2, \end{aligned}$$

which proves (4.10); here,  $\text{Var}_{i-1}$  denotes the conditional variance given  $H_{i-1}$ .

To prove (4.11), suppose in addition that the function  $g$  is convex in each of its arguments, as stated in the second part of Proposition 4.6. Let  $\tilde{\mathbb{E}}_i$  denote the conditional expectation given  $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$ . Then, for all  $i$ , by Jensen's inequality,

$$\begin{aligned} \mathbb{E}_{i-1} Y &= \mathbb{E}_{i-1} \tilde{\mathbb{E}}_i Y = \mathbb{E}_{i-1} \tilde{\mathbb{E}}_i g(X_1, \dots, X_n) \\ &\geq \mathbb{E}_{i-1} g(X_1, \dots, X_{i-1}, \tilde{\mathbb{E}}_i X_i, X_{i+1}, \dots, X_n) \\ &= \mathbb{E}_{i-1} g(X_1, \dots, X_{i-1}, \mathbb{E} X_i, X_{i+1}, \dots, X_n) = \mathbb{E}_{i-1} \tilde{Y}, \end{aligned}$$

in view of (5.15), if  $x_i$  is chosen to coincide with  $\mathbb{E} X_i$ ; hence,

$$\mathbb{E}_{i-1} \eta_i = \mathbb{E}_{i-1} Y - \mathbb{E}_{i-1} \tilde{Y} \geq 0.$$

This and formulas (5.17) and (5.18) imply that

$$\xi_i \leq \mathbb{E}_i \eta_i \leq \rho_i(X_i, \mathbb{E} X_i),$$

which is equivalent to (4.11).  $\square$

*Proof of Corollary 4.8.* This follows immediately from Theorem 4.5 and Proposition 4.6, with  $\rho_i(\tilde{x}_i, x_i) = \|\tilde{x}_i - x_i\|$ .  $\square$

## REFERENCES

- [1] BENTKUS, V. (2001). On measure concentration for separately Lipschitz functions in product spaces. *Israel J. Math.* To appear.
- [2] BENTKUS, V. (2002) A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand. *Lithuanian Math. J.* **42**, 262–269. MR 1947624
- [3] BENTKUS, V. (2003) An inequality for tail probabilities of martingales with differences bounded from one side. *J. Theoret. Probab.* **16**, 161–173. MR 1956826
- [4] BENTKUS, V. (2004) On Hoeffding's inequalities. *Ann. Probab.* **32**, 1650–1673. MR 2060313
- [5] BOBKOV, S. G., GÖTZE, F. AND HOUDRÉ, C. (2001) On Gaussian and Bernoulli covariance representations. *Bernoulli* **7**, 439–451. MR 1836739
- [6] EATON, M. L. (1970). A note on symmetric Bernoulli random variables. *Ann. Math. Statist.* **41**, 1223–1226. MR 268930
- [7] EATON, M. L. (1974). A probability inequality for linear combinations of bounded random variables. *Ann. Statist.* **2**, 609–614.
- [8] FIGIEL, T., HITCZENKO, P., JOHNSON, W. B., SCHECHTMAN, G. AND ZINN, J. (1997). Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.* **349**, 997–1027. MR 1390980

- [9] FUK, D. H. (1971). Certain probabilistic inequalities for martingales. *Siberian Math. J.* **14**, 131–137. MR 0293695
- [10] FUK, D. H. AND NAGAEV, S. V. (1971). Probabilistic inequalities for sums of independent random variables. (Russian. English summary) *Teor. Veroyatnost. i Primenen.* **16**, 660–675. MR 0293695
- [11] HAAGERUP, U. (1982). The best constants in the Khinchine inequality. *Studia Math.* **70**, 231–283. MR 0654838
- [12] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30. MR 144363
- [13] Hoeffding, W. (1955). The extrema of the expected value of a function of independent random variables. *Ann. Math. Statist.* **26**, 268–275. MR 70087
- [14] Hoeffding, W. AND Shrikhande, S. S. (1955). Bounds for the distribution function of a sum of independent, identically distributed random variables. *Ann. Math. Statist.* **26**, 439–449. MR 72377
- [15] KARLIN, S. AND STUDDEN, W. J. (1966). *Tchebycheff systems: With applications in analysis and statistics*. Pure and Applied Mathematics, Vol. XV. Interscience Publishers John Wiley & Sons, New York-London-Sydney. MR 204922
- [16] KARR, A. F. (1983). Extreme points of certain sets of probability measures, with applications. *Math. Oper. Res.* **8**, 1, 74–85. MR 703827
- [17] KHINCHIN, A. (1923). Über dyadische Brüche. *Math. Z.* **18**, 109–116.
- [18] MCDIARMID, C. (1989). On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*. London Math. Soc. Lecture Note Ser., Vol. **141**. Cambridge Univ. Press, Cambridge, 148–188. MR 1036755
- [19] MCDIARMID, C. (1998). Concentration. In *Probabilistic methods for algorithmic discrete mathematics*. Algorithms Combin., Vol. **16**. Springer, Berlin, 195–248. MR 1678578
- [20] NAGAEV, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7**, 745–789. MR 0542129
- [21] I. F. PINELIS (1981) Limit theorems on large deviations for sums of independent random variables with Cramer’s condition violated. (Russian) Deposited at VINITI (All-Russian Institute of Scientific and Technical Information or All-Union Institute of Scientific and Technical Information), No. 1674-81Dep., 94 pages.
- [22] PINELIS, I. F. (1985) Asymptotic equivalence of the probabilities of large deviations for sums and maximum of independent random variables. (Russian) *Limit theorems of probability theory, 144–173, 176, Trudy Inst. Mat.*, **5**, “Nauka” Sibirsk. Otdel., Novosibirsk. MR 0821760
- [23] PINELIS, I. (1994). Extremal probabilistic problems and Hotelling’s  $T^2$  test under a symmetry condition. *Ann. Statist.* **22**, 1, 357–368. MR 1272088
- [24] PINELIS, I. (1998). Optimal tail comparison based on comparison of moments. *High dimensional probability (Oberwolfach, 1996)*, 297–314, *Progr. Probab.*, **43**, Birkhäuser, Basel. MR 1652335
- [25] PINELIS, I. (1999). Fractional sums and integrals of  $r$ -concave tails and applications to comparison probability inequalities *Advances in stochastic inequalities (Atlanta, GA, 1997)*, 149–168, *Contemp. Math.*, **234**, Amer. Math. Soc., Providence, RI. MR 1694770
- [26] PINELIS, I. (2001). L’Hospital type rules for oscillation, with applications. *JIPAM. J. Inequal. Pure Appl. Math.* **2**, 3, Article 33, 24 pp. (electronic). MR 1876266
- [27] PINELIS, I. (2002). L’Hospital type results for monotonicity, with applications. *JIPAM. J. Inequal. Pure Appl. Math.* **3**, 1, Article 5, 5 pp. (electronic). MR 1888920
- [28] PINELIS, I. (2005). Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above. *Preprint*, <http://arxiv.org/abs/math.PR/0512301>.
- [29] PINELIS, I.; SAKHANENKO, A. I (1985). Remarks on inequalities for probabilities of large deviations. *Theory Probab. Appl.* **30**, 143–148. MR 0779438

- [30] SHORACK, G. R. AND WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York. MR 0838963
- [31] WHITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Teor. Veroyatnost. i Primenen.* **5**, 331–335. MR 0133849
- [32] YURINSKII , V. V. (JURINSKIĬ, V. V.) (1974). Exponential estimates for large deviations. (Russian) *Teor. Veroyatnost. i Primenen.* **19**, 152–154. MR 334298

DEPARTMENT OF MATHEMATICAL SCIENCES, MICHIGAN TECHNOLOGICAL UNIVERSITY, HOUGHTON, MICHIGAN 49931

*E-mail address:* `ipinelis@math.mtu.edu`